

# Isotopica: a tool for the calculation and viewing of complex isotopic envelopes

Jorge Fernandez-de-Cossio\*, Luis Javier Gonzalez, Yoshinori Satomi<sup>1</sup>, Lazaro Betancourt, Yassel Ramos, Vivian Huerta, Abel Amaro, Vladimir Besada, Gabriel Padron, Naoto Minamino<sup>2</sup> and Toshifumi Takao<sup>1</sup>

Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana, Cuba, <sup>1</sup>Institute for Protein Research, Osaka University, Yamadaoka 3-2, Suita, Osaka 565-0871, Japan and <sup>2</sup>Department of Pharmacology, National Cardiovascular Center Research Institute, Fujishirodai, Suita, Osaka 565-8565, Japan

Received February 20, 2004; Revised and Accepted April 8, 2004

## ABSTRACT

The web application Isotopica has been developed as an aid to the interpretation of ions that contain naturally occurring isotopes in a mass spectrum. It allows the calculation of mass values and isotopic distributions based on molecular formulas, peptides/proteins, DNA/RNA, carbohydrate sequences or combinations thereof. In addition, Isotopica takes modifications of the input molecule into consideration using a simple and flexible language as a straightforward extension of the molecular formula syntax. This function is especially useful for biomolecules, which are often subjected to additional modifications other than normal constituents, such as the frequently occurring post-translational modification in proteins. The isotopic distribution of any molecule thus defined can be calculated by considering full widths at half maximum or mass resolution. The combined envelope of several overlapping isotopic distributions of a mixture of molecules can be determined after specifying each molecule's relative abundance. The results can be displayed graphically on a local PC using the Isotopica viewer, a standalone application that is downloadable from the sites below, as a complement to the client browser. The  $m/z$  and intensity values can also be obtained in the form of a plain ASCII text file. The software has proved to be useful for peptide mass fingerprinting and validating an observed isotopic ion distribution with reference to the theoretical one, even from a multi-component sample. The web server can be

accessed at <http://bioinformatica.cigb.edu/isotopica> and <http://coco.protein.osaka-u.ac.jp/isotopica>.

## INTRODUCTION

Mass spectrometry (MS), an essential tool for proteomic analysis, allows the prompt identification of proteins in conjunction with a sequence database search. A mass spectrum consists of  $m/z$  values and intensities, where  $m$  and  $z$  denote mass and the number of charges on an ion, respectively. The  $m/z$  values are subjected to a database search, but the relative intensities of the ions, in general, do not correspond to the relative abundances of the analytes. However, relative quantification is possible when the analyte is compared with an isotopically labeled form of itself. The addition of such an internal standard entails no change in the ionization efficiency of the analyte (1). In addition, a mass spectrum of a natural compound gives the isotopic distribution, which is normally observed as the result of the presence of natural isotopes in the sample. Although the spacing between adjacent isotopic peaks and their relative abundance might be indicative of the charge states and the molecular formula of an ion, respectively, they can frequently interfere with a precise mass determination unless the isotopic peaks are taken into account or de-isotoped using software.

Many standalone software programs and some web application software programs are currently available for the calculation of theoretical isotopic distribution (e.g. MS-Isotope: <http://prospector.ucsf.edu/ucsfhtml4.0/msiso.htm>; Isotopident: <http://haven.isb-sib.ch/tools/isotopident/htdocs/>) (2) based on input sequences of biopolymers (proteins, DNA, sugars, etc.) or molecular formulas, and for *de novo*

\*To whom correspondence should be addressed. Email: [jorge.cossio@cigb.edu.cu](mailto:jorge.cossio@cigb.edu.cu)  
Correspondence may also be addressed to Toshifumi Takao. Email: [tak@protein.osaka-u.ac.jp](mailto:tak@protein.osaka-u.ac.jp)

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated.

sequencing (3–5) as well. Since the performance of MS has been significantly improved with respect to resolution and accuracy as the result of the development of the Fourier-transform ion cyclotron resonance (FTICR) mass spectrometer, complex and high-molecular-weight biopolymers such as glycoproteins can now be analyzed.

In order to support the calculation of the mass and theoretical isotopic distribution of a mixture of complex biopolymers, the web application Isotopica permits the flexible input of multiple components, based on protein sequence, molecular formulas, and so on. In addition, a standalone Windows application, downloadable from the main page, aids visualization of the calculated spectra on a local PC as a complement to the client browser. This application is useful for validating an observed peak alongside the theoretical one, especially one with a higher mass or a complex pattern due to the presence of more than two components or enriched stable isotopes within the isotopic distribution.

## Software

Isotopica is a .NET web application developed using the Microsoft Development Environment Visual Studio .NET version 7.0 (Copyright © 1987–2002, Microsoft Corporation). Isotopica is coded mainly in C++, C# and ASP.NET using the Microsoft .NET Framework Software Development Kit (SDK) (Copyright © 1998–2002, Microsoft Corporation). The Isotopica viewer was developed with Borland® Delphi™ Studio Enterprise version 7.0 (Copyright © 1983–2002, Borland Software Corporation).

Isotopica implements the algorithm proposed by Rockwood *et al.* (6) for the calculation of the isotopic distributions of the individual analytes. Briefly, let the isotopes of element *A* have masses  $m_{A1}$ ,  $m_{A2}$  with abundances  $p_{A1}$ ,  $p_{A2}$ , respectively, and the isotopes of element *B* have masses  $m_{B1}$ ,  $m_{B2}$  with abundances  $p_{B1}$ ,  $p_{B2}$ , respectively. The isotope abundance distribution of *A* can be represented by the summation of delta functions in the mass domain:  $D_A(m) = p_{A1}\delta(m - m_{A1}) + p_{A2}\delta(m - m_{A2})$ , and analogously for *B*:  $D_B(m) = p_{B1}\delta(m - m_{B1}) + p_{B2}\delta(m - m_{B2})$ . The isotopic distribution of a molecule with molecular formula *AB* can be obtained by convoluting the elements' isotopic distributions:  $D_{AB}(m) = \int D_A(m - x)D_B(x)dx$ . These principles are readily extended to any molecular formula by noting that for any other element *C*, the isotopic distribution of *ABC* can then be obtained after convoluting  $D_{AB}$  with  $D_C$ . Convolution in one domain corresponds to multiplication in the Fourier-transformed domain according to the convolution theorem. For a formula  $A_{nA}B_{nB}C_{nC}$ , the isotopic distribution can then be obtained by  $D_{A_{nA}B_{nB}C_{nC}} = F^{-1}[F(D_A)^{n_A}F(D_B)^{n_B}F(D_C)^{n_C}]$ , where  $F$  and  $F^{-1}$  represent the direct and inverse discrete Fourier transform respectively. The discrete Fourier transform is computed using the Fast Fourier Transforms (FFTs) algorithm (7). Composition of the components in a total envelope is obtained using the same sample intervals used for the FFT calculations.

## Input

Isotopica allows the input of a mixture of sequences of peptides/proteins (Figure 1), DNA/RNA, carbohydrates and molecular formulas. Modifications to the registered molecules can

be specified using the extended formula syntax described below.

## Extended use of the molecular formula syntax

Natural element symbols are usually spelled starting with an uppercase character, followed by lowercase characters. Since the conventional three-letter-code symbols for amino acids start with uppercase, followed by lowercase, even a mixture of one- and three-letter codes in the same sequence can be unambiguously deciphered. Codes with more than one letter for typical and rare modifications as well as artificial amino acids can be used together with the compact one-letter-code sequence of standard amino acids without explicit specification of the code length currently in use. For example, the peptide sequences ALHPY and ALeuHProY are equivalently deciphered. The sequence ATRDCamY readily highlights the modified amino acid as carbamidomethylcysteine (Cam), which can be registered in 'Residue registration', without the need to switch to a three-letter code for the rest of the amino acids. The advantage of this extension becomes more apparent when uncommon modified amino acids are dealt with. The same formatting is applicable to nucleic acids. Since the nomenclature for carbohydrates does not follow strictly the rule that only the first letter is capitalized, e.g. GlcNAc and NeuAc, only space-delimited symbols are accepted (see the lower panel of Figure 2).

A modification usually entails the loss or the incorporation of groups in a molecule. In order to support both events in the same formula, the molecular parser of Isotopica also allows for the indication of negative numbers as a subscript of the elements or molecular units involved in the loss. This same extension is considered for amino acids, nucleic acids, and carbohydrates. For example, deamidation, which can often take place in peptides and proteins, can be specified by inputting the original amino acid sequences followed by the formula, '[NH2]–1 [OH]' (see Figure 3).

The molecular context of the 'formula' in each option is provided by comma-delimited ordering and combo-box selection (Figures 2 and 3) to unambiguously differentiate different compounds using the same coding, e.g. H for hydrogen from H for histidine, and A for alanine from A for adenosine.

## Peptide mass fingerprinting (Isotopica Digest)

Multiple protein sequences are input in the FASTA format. Peptide sets are generated according to the enzyme used for each sequence (Figure 1). Similarly to other software, Isotopica allows various settings for an enzyme to be used by the number of missing cleavages, charge states, mono-isotopic or average mass and N- and C-terminal modifications that can be set by a molecular formula. 'Peptide filter' and '*m/z* filter' allow the selection of peptides which have a specific amino acid or sequence and whose masses are in a given *m/z* range. In the output, the molecular formulas, molecular weights and *m/z* values of each constituent peptide of the digests of proteins are generated, and their isotopic distributions are obtained by Isotopica Simulator (see below), which can be launched by clicking molecular formulas or sequences.

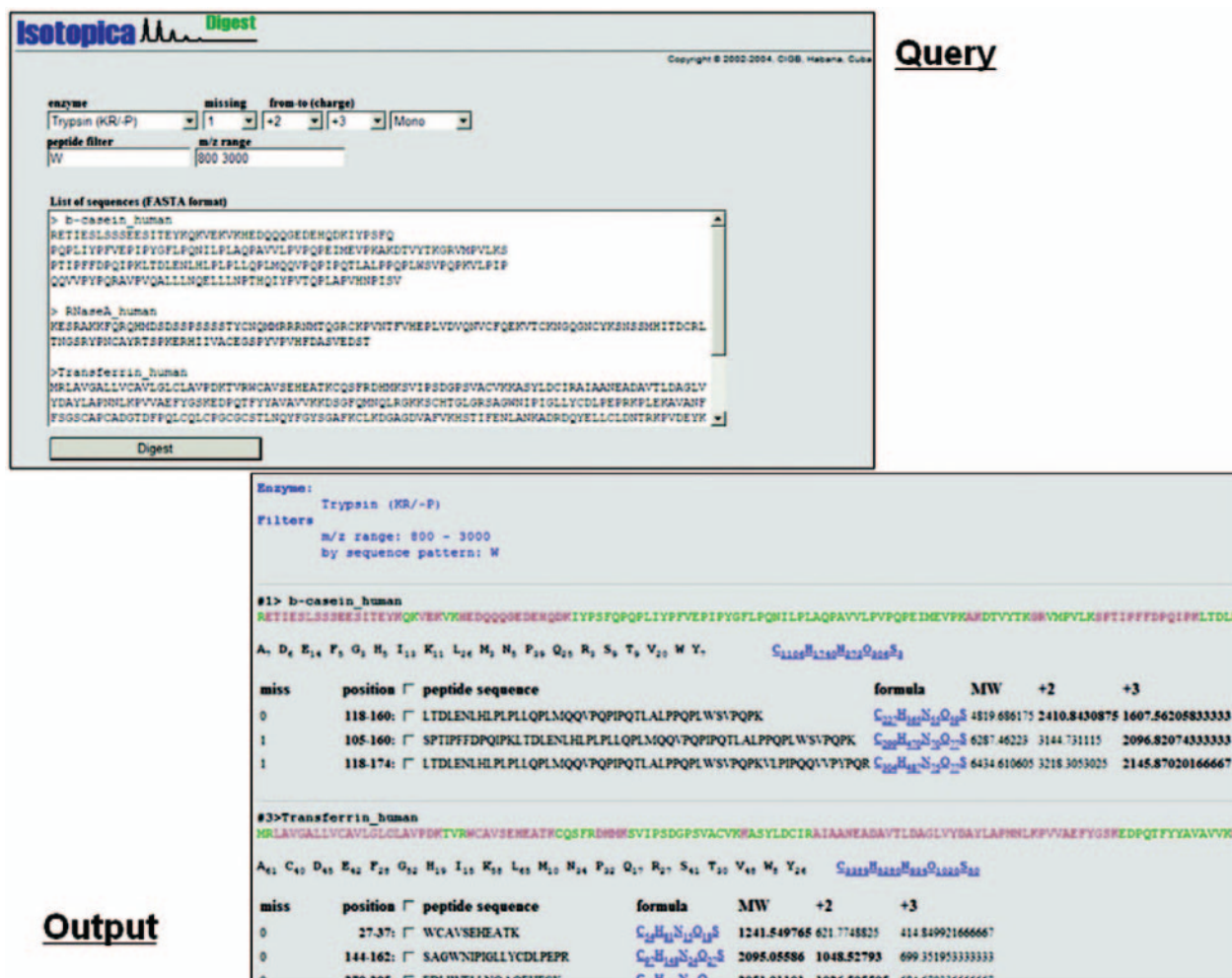


Figure 1. Typical input and output formats for peptide mass fingerprinting.

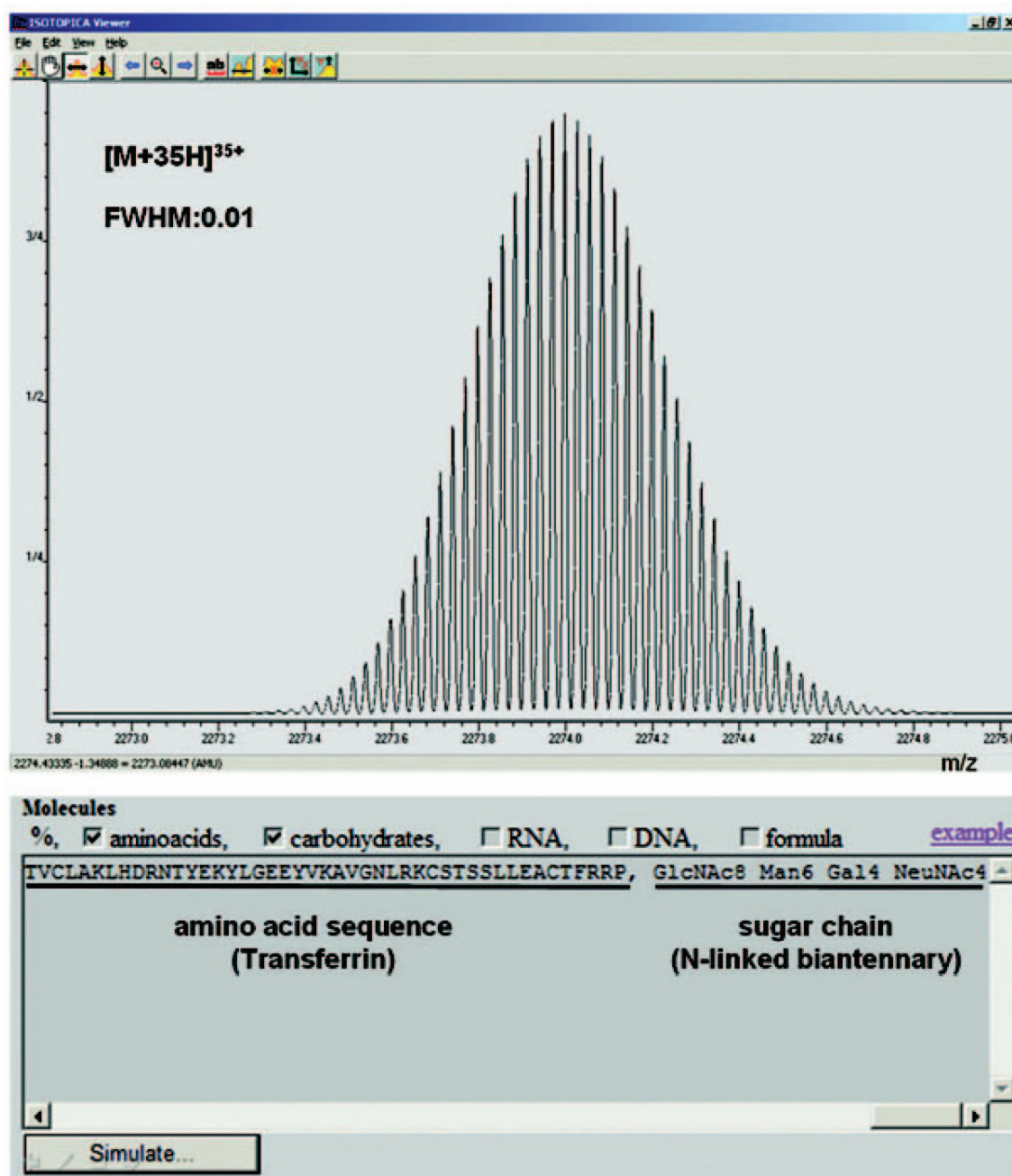
### Viewing isotopic distribution (Isotopica Simulator)

The theoretical isotopic distribution is calculated by the server, and the results can be displayed graphically on a local PC using the Isotopica viewer, a standalone application downloadable from the home page, as a complement to the client browser.

The Isotopica viewer allows the graphical reconstitution of theoretical isotopic peaks of an ion in terms of full widths at half maximum (FWHM) or mass resolution, the charge state of an ion, artificial shifts in mass and center mass for display range within 50 Da, all of which can be set by the individual user. Figure 2 shows a typical example of output for a large complex molecule which comprises 679 amino acids and 22 sugar units. The isotopic envelope for a 35<sup>+</sup>-charged ion, which could be observed using an FTICR-MS equipped with an electrospray ion source (8), was obtained using a value of 0.01 for FWHM with well-resolved isotopic peaks.

Isotopica can also be used to calculate the isotopic distributions of multiple components based on the relative abundance

of each component, as set by the user, and to integrate them into the new isotopic envelope. In addition, the Isotopica viewer allows the user to copy and paste a raw spectrum as ASCII text formatted as a list of *m/z* values and intensities in two columns separated by a space, for comparison with a calculated theoretical isotopic distribution. Figure 3 shows a comparison between the raw MS spectrum of a 31-amino acid peptide and the theoretical isotopic distributions calculated from the sequence. Since this sample was obtained experimentally as an equal mixture of  $\beta$ -endorphin (*m/z* 3463.8) and its deamidated form (*m/z* 3464.8), their isotopic envelopes were estimated to overlap within 1 Da. The isotopic distributions were then calculated for the mixtures of these peptides with the relative ratios of 1:0, 4:1, 1:1, and 1:4 ( $\beta$ -endorphin:the deamidated form) and compared with the observed distributions. As a result, the isotopic envelope of a 1:1 mixture turned out to coincide with the observed one, demonstrating the great advantage of Isotopica for fine comparison between observed and theoretical isotopic envelopes. This function can be useful not only for validating the purity of an analyte compound, but



**Figure 2.** Output from the Isotopica viewer of  $35^+$ -charged ion of transferrin [MW (average): 79 555.559], a 679-amino acid glycoprotein with two bi-antennary carbohydrates (GlcNAc2 Man3 GlcNAc2 Gal2 NeuAc2). The lower panel is a screen dump of the input from where the sequence of amino acids (one-letter code) and the composition of carbohydrates are typed in.

also for constructing an *in silico* MS spectrum based on the given abundances of each component in a mixture.

## ACKNOWLEDGEMENTS

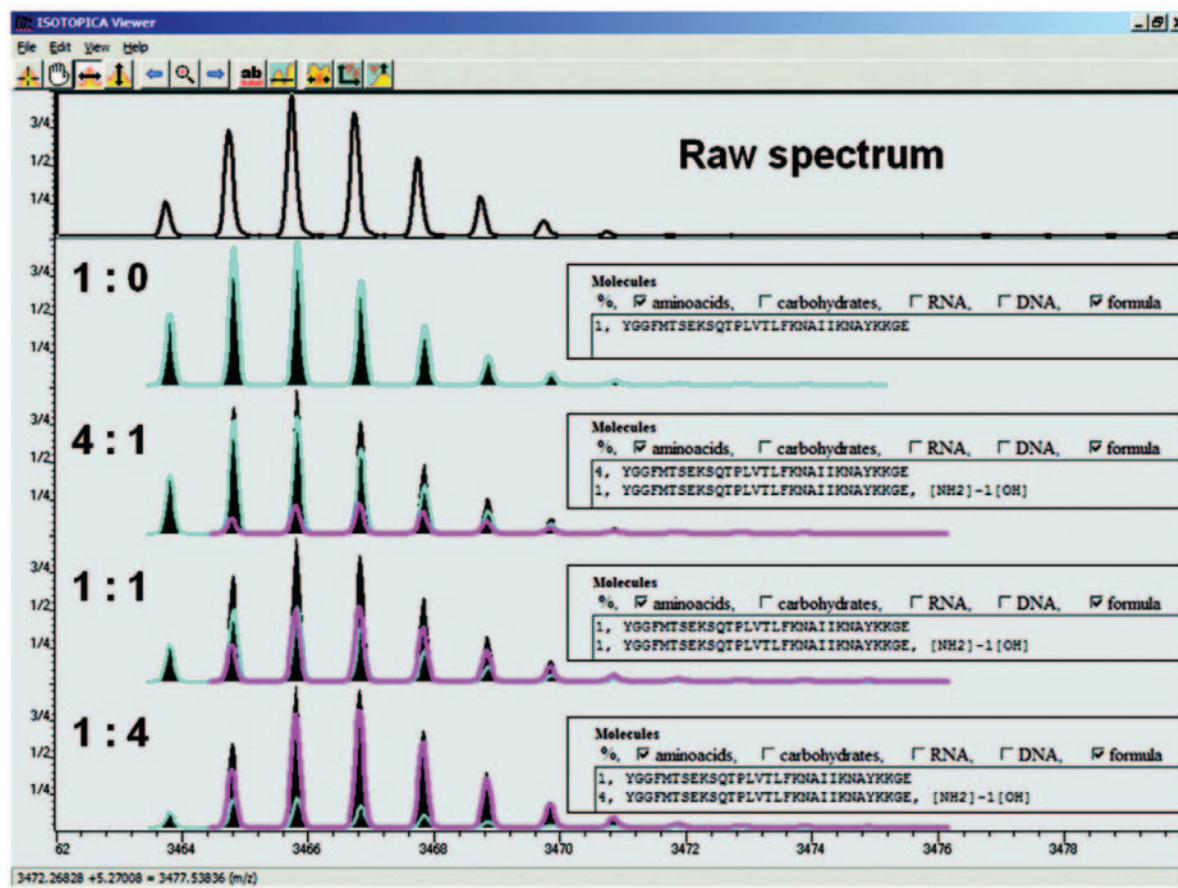
This study was supported by the research and development funds of the Center for Genetic Engineering and Biotechnology, the Special Coordination Fund for the Promotion of Science and Technology (Intellectual Infrastructure Research

and Development for Its Comprehensive Promotion System) and Grants-in-Aid for Scientific Research (No. 14380293 to T.T.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

- Gobom, J., Krauter, K.O., Persson, R., Steen, H., Roepstorff, P. and Ekman, R. (2000) Detection and quantification of neurotensin in human





**Figure 3.** Comparison between observed (black open peaks in the top panel) and the sum of the calculated theoretical isotopic distributions (baseline-filled peaks in the lower panels) computed by Isotopica for the mixtures of peptides YGGFMTSEKSTPLVTLFKNAIIKNAYKKGE (turquoise trace) and its deamidated form (Asn20→Asp) (pink trace) in the relative ratios 1:0, 4:1, 1:1, and 1:4. The inset boxes are screen dumps of each input for calculation, where the relative ratios of the two components, their amino acid sequences and the formula for modification are typed in.

- brain tissue by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.*, **72**, 3320–3326.
- Yergey, J.A. (1987) A general approach to calculating isotopic distributions for mass spectrometry. *Int. J. Mass Spectrom. Ion Phys.*, **52**, 337–349.
- Taylor, J.A. and Johnson, R.S. (1997) Sequence database searches via *de novo* peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, **11**, 1067–1075.
- F.-Cossio, J., Gonzalez, J., Betancourt, L., Besada, V., Padron, G., Shimonishi, Y. and Takao, T. (1998) Automated interpretation of high-energy collision-induced dissociation spectra of singly protonated peptides by 'SeqMS', a software aid for *de novo* sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, **12**, 1867–1878.
- Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. and Lajoie, G. (2003) PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, **17**, 2337–2342.
- Rockwood, A.L., Van Orden, S.L. and Smith, R.D. (1995) Rapid calculation of isotope distributions. *Anal. Chem.*, **67**, 2699–2704.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1997) *Numerical recipes in C. The Art of Scientific Computing*, 2nd edn. Cambridge University Press, Cambridge.
- Senko, M.W., Hendrickson, C.L., Pasa-Tolic, L., Marto, J.A., White, F.M., Guan, S. and Marshall, A.G. (1996) Electrospray ionization Fourier transform ion cyclotron resonance at 9.4 T. *Rapid Commun. Mass Spectrom.*, **10**, 1824–1828.